

# UC Berkeley

## Proposals from the Script Encoding Initiative

### Title

Final proposal for encoding the Warang Citi script in the SMP of the UCS

### Permalink

<https://escholarship.org/uc/item/2wm3g87z>

### Author

Everson, Michael

### Publication Date

2012-04-19

Peer reviewed

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Final proposal for encoding the Warang Citi script in the SMP of the UCS****Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)****Author: Michael Everson****Status: Liaison Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N1958 (1999-01-29), N3411 (2008-04-08), N3668 (2009-08-05)****Date: 2012-04-19**

**1. Introduction.** The Warang Citi script is used to write the Ho language. Ho is a North Munda language, which family, together with the Mon-Khmer languages, makes up Austro-Asiatic. Warang Citi was devised by charismatic community leader Lako Bodra as part of a comprehensive cultural program, and was offered as an improvement over scripts used by Christian missionary linguists. Ho people live in the Indian states of Orissa and Jharkand. In Jharkand, they are found in Ranchi, Chaibasa and Jamshedpur, and in villages like Pardsa, Jaldhar, Tekasi, Tilupada, Baduri, Purtydigua, Roladih, Tupu Dana, Jetia, Dumbisai, Harira, Gitilpi, Karlajuri, Sarliburu, Narsana, Gidibas Kokcho, Lupungutu, Pandaveer, Jhinkapani, Kondwa. According to the SIL *Ethnologue*, there are 1,026,000 speakers of Ho. There are at present two publications in the script: a magazine *Ho Sanagam* ('meeting' from Hindi *saṅgam*), which comes out yearly and *Kolhan Sakam*, which comes out biweekly.

Today, the Ho community can be characterized as still a primarily oral community, with an emergent literary tradition. Many Ho do not write their language in any form. Among those who do, there are different camps. One camp that promotes the use of Devanagari for Ho; another camp knows how to use Warang Citi, and expresses a strong sentiment, even a kind of reverence towards it. And there are some people who seem to fall into both camps. There are also people who use Latin letters to write Ho on an ad-hoc basis and for things like e-mail or SMS. The use of Warang Citi seems to be stronger in the eastern part of the diaspora, where books that use it have been printed. (See samples below.) The expert we consulted, Mr Dobro Buriuli, teaches the script and over 200 students have passed his courses. In the opinion of Anna Pucilowski, who did fieldwork in February 2012, "the people that are interested in Warang Chiti are all ages I would say but they're mostly educated. Literacy is pretty low outside the towns and cities."

**2. Structure.** Warang Citi is a casing script, written from left to right. Unlike many other Indic scripts, it has features of an alphabet: vowels are written as full letters, with no vowel-modifiers. However, consonants may have an inherent vowel; it typically is pronounced [a] or [ɔ], and less often [ɛ], and never occurs in final position. Because the inherent vowels do not appear explicitly, there can be ambiguity in the reading of certain words. The script has no regular system of conjuncts nor an explicit virama. In the presentation here, the inherent vowel is transcribed *Aa* and the explicit vowel ᱠᱣ is transcribed *Āā*.

The letter ᱠᱣ ᱠᱣ HIYO is used after a vowel to indicate lengthening or aspiration: ᱠᱣ ᱠᱣ ᱠᱣ ᱠᱣ (Ā ā). Long *i* and *u* are represented by the individual letters ᱠᱣ ᱠᱣ *Ī ī* and ᱠᱣ ᱠᱣ *Ū ū*; these do not appear to be equivalent to ᱠᱣ ᱠᱣ *IH ih* or ᱠᱣ ᱠᱣ *UH uh*. The letters ᱠᱣ ᱠᱣ *Ī ī* and ᱠᱣ ᱠᱣ *Ū ū* may also, in certain cases, be pronounced as short vowels or diphthongs ([ai] for long *i* or [au] for long *u*). The details of how (or if) length is handled orthographically in Warang Citi is unclear at this point.

As in the Brahmi-derived scripts, Warang Citi includes the OM sign  $\phi$ , which occasionally appears in words, such as  $\phi\phi\phi$  *šoms*, representing Hindi *śvāms* ‘breathing’. The script does not include a diacritic mark for anusvara as in Devanagari, but rather has a separate grapheme,  $\nu$  NGAA.

The script has four so-called “ligatures”  $\Upsilon$  YA ( $V + V_M + M$ ),  $\mathbb{F}$  YO ( $\Gamma + \Gamma_A + A$ ),  $\mathbb{E}$  II ( $\mathbb{F}_I + \mathbb{F}_I$ ), and  $\mathbb{U}$  UU ( $L_U + L_U$ ). None of these “ligatures” are decomposable, and so are encoded separately. The digraph  $hb$  occurs occasionally and is used for [b] or [w].

**3. Digits and numbers.** Warang Citi now makes use of a zero, though this may not have been the case previously; no zero is given in Pinnow 1972 or Zide 1996, but more recent materials use zero, as for instance in the date  $\mathbb{J}\mathbb{W}-\mathbb{J}\mathbb{O}-\mathbb{A}\mathbb{O}\mathbb{O}\mathbb{M}$  ‘15-10-2003’ (see Figure 6). Note that the higher digits may have fallen into disuse. In the date, for example, 118EA WARANG CITI NUMBER TEN is not used, but rather 118E1 WARANG CITI DIGIT ONE and 118E0 WARANG CITI DIGIT ZERO. No symbols for 100 or 1000 are known, though Pinnow postulated that they may exist.

**4. Sorting.** Collation order is as in the code chart, with casing pairs interfiling as in Latin. An exception here is  $\mathbb{U}+118\mathbb{F}\mathbb{F}$  WARANG CITI OM, which does not file like  $\mathbb{Z}\mathbb{V}\mathbb{z}\mathbb{v}$  OM om, but rather comes at the beginning of the alphabet, although it is placed last in the code chart so that the alphabet itself can have a constant offset. The order is as given in Figure 5, which is considered authoritative. In the code chart all the capital letters are given, then all the small letters, as in ASCII, Armenian, Cyrillic, Deseret, Glagolitic, and Greek.

$\phi$  om <  $V_M$  <<<  $\nu m$  <  $\Gamma \mathring{A}$  <<<  $\mathring{s} \mathring{a}$  <  $\mathbb{F} I$  <<<  $\mathring{f} i$  <  $L_U$  <<<  $\mathring{l} u$  <  $\Upsilon P$  <<<  $\mathring{y} \mathring{p}$  <  $\mathbb{F} Y$  <<<  $\mathring{y} y$  <  
 $\mathbb{E} \mathring{I}$  <<<  $\mathring{r} \mathring{i}$  <  $\mathbb{U} \mathring{U}$  <<<  $\mathring{r} \mathring{u}$  <  $\nabla E$  <<<  $\mathring{v} e$  <  $\mathbb{Z} O$  <<<  $\mathring{z} o$  <  $\mathbb{J} \mathring{N}$  <<<  $\mathring{z} \mathring{n}$  <  $\mathbb{A} G$  <<<  $\mathring{s} g$  <  
 $\mathbb{K} K$  <<<  $\mathring{s} k$  <  $\mathbb{C} \mathring{N}$  <<<  $\mathring{s} \mathring{n}$  <  $\mathbb{E} J$  <<<  $\mathring{e} j$  <  $\mathbb{C} C$  <<<  $\mathring{s} c$  <  $\mathbb{N} N$  <<<  $\mathring{s} n$  <  $\mathring{r} D$  <<<  $\mathring{r} d$  <  
 $L_T$  <<<  $\mathring{l} t$  <  $\mathbb{N} N$  <<<  $\mathring{n} n$  <  $\mathbb{U} D$  <<<  $\mathring{v} d$  <  $\mathbb{O} T$  <<<  $\mathring{t} t$  <  $\mathbb{A} M$  <<<  $\mathring{s} m$  <  $\mathbb{D} B$  <<<  $\mathring{b} b$  <  
 $\mathbb{U} P$  <<<  $\mathring{v} p$  <  $\mathbb{U} H$  <<<  $\mathring{v} h$  <  $\mathbb{D} L$  <<<  $\mathring{v} l$  <  $\mathbb{S} R$  <<<  $\mathring{s} r$  <  $\mathbb{T} R$  <<<  $\mathring{s} r$  <  $\mathbb{U} \mathring{S}$  <<<  $\mathring{s} s$  <  
 $\mathbb{V} S$  <<<  $\mathring{s} s$  <  $\mathbb{H} V$  <<<  $\mathring{s} v$

**5. Character names.** The name of the script is  $\mathbb{U}\mathring{o}\mathring{s}\mathring{r}\mathring{z}\mathring{g}\mathring{f}\mathring{g}\mathring{f}$  *Hbârṇ citi* ‘undying writing’, where  $\mathring{v}\mathring{o}$  *hb* is a digraph representing /w/. The name *Varang Kshiti* found in some literature is a sanskritized form. Letter names are traditional. Pinnow 1972 gives the following transcriptions. After those are given the recommended spellings for English and French. The name  $\mathbb{V}\mathbb{I}\mathbb{Y}\mathbb{O}$  for the last letter  $\mathbb{H}\mathbb{S}$  [v] has been constructed; it is an original ligature of  $\mathbb{J}\mathring{s}$  A and  $\mathbb{U}\mathring{v}$  HIYO.

$\eta\mathring{a}$ ?	NGAA	$\mathring{N}\mathring{A}$ ’	$\mathring{?}\mathring{e}$ ?	E	’E’	$\mathring{?}\mathring{e}$ ?	ENN	’ENN	$\mathring{p}\mathring{u}$ ?	PU	PU’
$\mathring{?}\mathring{a}$	A	’A	$\mathring{?}\mathring{o}$ ?	O	’O’	$\mathring{?}\mathring{o}$ ?	ODD	’ODD	$\mathring{h}\mathring{i}\mathring{y}\mathring{o}$	HIYO	HIYO
$\mathring{w}\mathring{ü}$ ?	WI	$\mathring{Ü}$ ’I	$\mathring{?}\mathring{a}\mathring{ng}$	ANG	’ANG	$\mathring{t}\mathring{e}$ ?	TTE	TTE’	$\mathring{h}\mathring{o}\mathring{l}\mathring{o}$	HOLO	HOLO
$\mathring{y}\mathring{u}$ ?	YU	$\mathring{YU}$ ’U	$\mathring{g}\mathring{a}$ ?	GA	GA’	$\mathring{n}\mathring{u}\mathring{ṅ}$	NUNG	NUNG	$\mathring{h}\mathring{o}\mathring{r}$	HORR	HORR
$\mathring{y}\mathring{a}$ ?	YA	YA’	$\mathring{k}\mathring{o}$ ?	KO	KO’	$\mathring{d}\mathring{a}$ ?	DA	DA’	$\mathring{h}\mathring{a}\mathring{r}$	HAR	HAR
$\mathring{y}\mathring{o}$ ?	YO	YO’	$\mathring{?}\mathring{e}\mathring{n}$	ENY	’EÑ	$\mathring{?}\mathring{a}\mathring{t}$	AT	’AT	$\mathring{s}\mathring{u}$ :	SSUU	$\mathring{Ç}\mathring{U}$
$\mathring{?}\mathring{i}$ :	II	’I	$\mathring{y}\mathring{ü}$ ’j	YUJ	YÜ’J	$\mathring{?}\mathring{a}\mathring{m}$	AM	’AM	$\mathring{s}\mathring{i}$ :	SII	$\mathring{S}\mathring{I}$
$\mathring{?}\mathring{u}$ :	UU	’U	$\mathring{?}\mathring{ü}\mathring{c}$	UC	’ÜC	$\mathring{b}\mathring{u}$	BU	BU	$\mathring{v}\mathring{i}\mathring{y}\mathring{o}$	VIYO	VIYO

**6. Punctuation.** Pinnow states explicitly that European punctuation is used as in English. The following characters have been observed: . , : ... ‘ ’ “ ” - — ? and !. In some fonts the dots tend to be drawn rather high—but this is still FULL STOP, not MIDDLE DOT. The apostrophe ’ (RIGHT SINGLE QUOTATION MARK) is used very frequently in words ending in -’ $\mathring{y}$  -’ $\mathring{z}$ , as in  $\mathring{r}\mathring{s}\mathring{e}\mathring{s}$ ’ $\mathring{y}$  *râjâ*’ $\mathring{z}$ .

**7. Conjunct forms.** A small number of conjunct forms are known, most of these represent doubled consonants. Pinnow calls them “ligatures” and says:

‘Other ligatures do not appear at all in the older books; in later ones ligatures were occasionally used, especially for the transliteration of Hindi words. The ligatures consist of the simple positioning of the second of two consonants which are not separated by a vowel on top of the first, e.g.  $b(a)r^n$ , actually  $br^n$ ... Since the ligatures occur only rarely, printing and writing the alphabet are relatively simple.’

[illegible]

**8. Linebreaking.** Letters and digits behave as in Latin and other alphabetic scripts. U+118FF ᮚ WARANG CITI OM behaves like a letter.





















































































118A0;	WARANG	CITI	CAPITAL	LETTER	NGAA;Lu;0;L;;;N;;;118C0;
118A1;	WARANG	CITI	CAPITAL	LETTER	A;Lu;0;L;;;N;;;118C1;
118A2;	WARANG	CITI	CAPITAL	LETTER	WI;Lu;0;L;;;N;;;118C2;
118A3;	WARANG	CITI	CAPITAL	LETTER	YU;Lu;0;L;;;N;;;118C3;
118A4;	WARANG	CITI	CAPITAL	LETTER	YA;Lu;0;L;;;N;;;118C4;
118A5;	WARANG	CITI	CAPITAL	LETTER	YO;Lu;0;L;;;N;;;118C5;
118A6;	WARANG	CITI	CAPITAL	LETTER	II;Lu;0;L;;;N;;;118C6;
118A7;	WARANG	CITI	CAPITAL	LETTER	UU;Lu;0;L;;;N;;;118C7;
118A8;	WARANG	CITI	CAPITAL	LETTER	E;Lu;0;L;;;N;;;118C8;
118A9;	WARANG	CITI	CAPITAL	LETTER	O;Lu;0;L;;;N;;;118C9;
118AA;	WARANG	CITI	CAPITAL	LETTER	ANG;Lu;0;L;;;N;;;118CA;
118AB;	WARANG	CITI	CAPITAL	LETTER	GA;Lu;0;L;;;N;;;118CB;
118AC;	WARANG	CITI	CAPITAL	LETTER	KO;Lu;0;L;;;N;;;118CC;
118AD;	WARANG	CITI	CAPITAL	LETTER	ENY;Lu;0;L;;;N;;;118CD;
118AE;	WARANG	CITI	CAPITAL	LETTER	YUJ;Lu;0;L;;;N;;;118CE;
118AF;	WARANG	CITI	CAPITAL	LETTER	UC;Lu;0;L;;;N;;;118CF;
118B0;	WARANG	CITI	CAPITAL	LETTER	ENN;Lu;0;L;;;N;;;118D0;
118B1;	WARANG	CITI	CAPITAL	LETTER	ODD;Lu;0;L;;;N;;;118D1;
118B2;	WARANG	CITI	CAPITAL	LETTER	TTE;Lu;0;L;;;N;;;118D2;
118B3;	WARANG	CITI	CAPITAL	LETTER	NUNG;Lu;0;L;;;N;;;118D3;
118B4;	WARANG	CITI	CAPITAL	LETTER	DA;Lu;0;L;;;N;;;118D4;
118B5;	WARANG	CITI	CAPITAL	LETTER	AT;Lu;0;L;;;N;;;118D5;

118B6;WARANG CITI CAPITAL LETTER AM;Lu;0;L;;;N;;;118D6;  
 118B7;WARANG CITI CAPITAL LETTER BU;Lu;0;L;;;N;;;118D7;  
 118B8;WARANG CITI CAPITAL LETTER PU;Lu;0;L;;;N;;;118D8;  
 118B9;WARANG CITI CAPITAL LETTER HIYO;Lu;0;L;;;N;;;118D9;  
 118BA;WARANG CITI CAPITAL LETTER HOLO;Lu;0;L;;;N;;;118DA;  
 118BB;WARANG CITI CAPITAL LETTER HORR;Lu;0;L;;;N;;;118DB;  
 118BC;WARANG CITI CAPITAL LETTER HAR;Lu;0;L;;;N;;;118DC;  
 118BD;WARANG CITI CAPITAL LETTER SSUU;Lu;0;L;;;N;;;118DD;  
 118BE;WARANG CITI CAPITAL LETTER SII;Lu;0;L;;;N;;;118DE;  
 118BF;WARANG CITI CAPITAL LETTER VIYO;Lu;0;L;;;N;;;118DF;  
 118C0;WARANG CITI SMALL LETTER NGAA;Ll;0;L;;;N;;;118A0;;118A0  
 118C1;WARANG CITI SMALL LETTER A;Ll;0;L;;;N;;;118A1;;118A1  
 118C2;WARANG CITI SMALL LETTER WI;Ll;0;L;;;N;;;118A2;;118A2  
 118C3;WARANG CITI SMALL LETTER YU;Ll;0;L;;;N;;;118A3;;118A3  
 118C4;WARANG CITI SMALL LETTER YA;Ll;0;L;;;N;;;118A4;;118A4  
 118C5;WARANG CITI SMALL LETTER YO;Ll;0;L;;;N;;;118A5;;118A5  
 118C6;WARANG CITI SMALL LETTER II;Ll;0;L;;;N;;;118A6;;118A6  
 118C7;WARANG CITI SMALL LETTER UU;Ll;0;L;;;N;;;118A7;;118A7  
 118C8;WARANG CITI SMALL LETTER E;Ll;0;L;;;N;;;118A8;;118A8  
 118C9;WARANG CITI SMALL LETTER O;Ll;0;L;;;N;;;118A9;;118A9  
 118CA;WARANG CITI SMALL LETTER ANG;Ll;0;L;;;N;;;118AA;;118AA  
 118CB;WARANG CITI SMALL LETTER GA;Ll;0;L;;;N;;;118AB;;118AB  
 118CC;WARANG CITI SMALL LETTER KO;Ll;0;L;;;N;;;118AC;;118AC  
 118CD;WARANG CITI SMALL LETTER ENY;Ll;0;L;;;N;;;118AD;;118AD  
 118CE;WARANG CITI SMALL LETTER YUJ;Ll;0;L;;;N;;;118AE;;118AE  
 118CF;WARANG CITI SMALL LETTER UC;Ll;0;L;;;N;;;118AF;;118AF  
 118D0;WARANG CITI SMALL LETTER ENN;Ll;0;L;;;N;;;118B0;;118B0  
 118D1;WARANG CITI SMALL LETTER ODD;Ll;0;L;;;N;;;118B1;;118B1  
 118D2;WARANG CITI SMALL LETTER TTE;Ll;0;L;;;N;;;118B2;;118B2  
 118D3;WARANG CITI SMALL LETTER NUNG;Ll;0;L;;;N;;;118B3;;118B3  
 118D4;WARANG CITI SMALL LETTER DA;Ll;0;L;;;N;;;118B4;;118B4  
 118D5;WARANG CITI SMALL LETTER AT;Ll;0;L;;;N;;;118B5;;118B5  
 118D6;WARANG CITI SMALL LETTER AM;Ll;0;L;;;N;;;118B6;;118B6  
 118D7;WARANG CITI SMALL LETTER BU;Ll;0;L;;;N;;;118B7;;118B7  
 118D8;WARANG CITI SMALL LETTER PU;Ll;0;L;;;N;;;118B8;;118B8  
 118D9;WARANG CITI SMALL LETTER HIYO;Ll;0;L;;;N;;;118B9;;118B9  
 118DA;WARANG CITI SMALL LETTER HOLO;Ll;0;L;;;N;;;118BA;;118BA  
 118DB;WARANG CITI SMALL LETTER HORR;Ll;0;L;;;N;;;118BB;;118BB  
 118DC;WARANG CITI SMALL LETTER HAR;Ll;0;L;;;N;;;118BC;;118BC  
 118DD;WARANG CITI SMALL LETTER SSUU;Ll;0;L;;;N;;;118BD;;118BD  
 118DE;WARANG CITI SMALL LETTER SII;Ll;0;L;;;N;;;118BE;;118BE  
 118DF;WARANG CITI SMALL LETTER VIYO;Ll;0;L;;;N;;;118BF;;118BF  
 118E0;WARANG CITI DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;  
 118E1;WARANG CITI DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;  
 118E2;WARANG CITI DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;  
 118E3;WARANG CITI DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;  
 118E4;WARANG CITI DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;  
 118E5;WARANG CITI DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;  
 118E6;WARANG CITI DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;  
 118E7;WARANG CITI DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;  
 118E8;WARANG CITI DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;  
 118E9;WARANG CITI DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;  
 118EA;WARANG CITI NUMBER TEN;No;0;L;;;10;N;;;;;  
 118EB;WARANG CITI NUMBER TWENTY;No;0;L;;;20;N;;;;;  
 118EC;WARANG CITI NUMBER THIRTY;No;0;L;;;30;N;;;;;  
 118ED;WARANG CITI NUMBER FORTY;No;0;L;;;40;N;;;;;  
 118EE;WARANG CITI NUMBER FIFTY;No;0;L;;;50;N;;;;;  
 118EF;WARANG CITI NUMBER SIXTY;No;0;L;;;60;N;;;;;  
 118F0;WARANG CITI NUMBER SEVENTY;No;0;L;;;70;N;;;;;  
 118F1;WARANG CITI NUMBER EIGHTY;No;0;L;;;80;N;;;;;  
 118F2;WARANG CITI NUMBER NINETY;No;0;L;;;90;N;;;;;  
 118FF;WARANG CITI OM;Lo;0;L;;;N;;;;;

## 10. Bibliography.

- Pinnow, Heinz-Jürgen. 1972. “Schrift und Sprache in den Werken Lako Bodras im Gebiet der Ho von Singbhum (Bihar)”, in *Anthropos* 67:822–857.
- Zide, Norman. 1996. “Scripts for Munda languages”, in Peter T. Daniels and William Bright, eds. *The world’s writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- [s.n.]. 2003. *ᱥᱟᱱ ᱦᱚᱱᱚ* (*Şişu halam*). [s.l.]: [s.n.]. (See Figure 25.)

**11. Acknowledgements.** This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Warang Citi encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	118A	118B	118C	118D	118E	118F
0	 118A0	 118B0	 118C0	 118D0	 118E0	 118F0
1	 118A1	 118B1	 118C1	 118D1	 118E1	 118F1
2	 118A2	 118B2	 118C2	 118D2	 118E2	 118F2
3	 118A3	 118B3	 118C3	 118D3	 118E3	
4	 118A4	 118B4	 118C4	 118D4	 118E4	
5	 118A5	 118B5	 118C5	 118D5	 118E5	
6	 118A6	 118B6	 118C6	 118D6	 118E6	
7	 118A7	 118B7	 118C7	 118D7	 118E7	
8	 118A8	 118B8	 118C8	 118D8	 118E8	
9	 118A9	 118B9	 118C9	 118D9	 118E9	
A	 118AA	 118BA	 118CA	 118DA	 118EA	
B	 118AB	 118BB	 118CB	 118DB	 118EB	
C	 118AC	 118BC	 118CC	 118DC	 118EC	
D	 118AD	 118BD	 118CD	 118DD	 118ED	
E	 118AE	 118BE	 118CE	 118DE	 118EE	
F	 118AF	 118BF	 118CF	 118DF	 118EF	 118FF

**Uppercase vowels**

118A0	᱆	WARANG CITI CAPITAL LETTER NGAA
118A1	᱇	WARANG CITI CAPITAL LETTER A
118A2	᱈	WARANG CITI CAPITAL LETTER WI
118A3	᱉	WARANG CITI CAPITAL LETTER YU
118A4	᱊	WARANG CITI CAPITAL LETTER YA
118A5	᱋	WARANG CITI CAPITAL LETTER YO
118A6	᱌	WARANG CITI CAPITAL LETTER II
118A7	ᱍ	WARANG CITI CAPITAL LETTER UU
118A8	ᱎ	WARANG CITI CAPITAL LETTER E
118A9	ᱏ	WARANG CITI CAPITAL LETTER O

**Uppercase consonants**

118AA	᱐	WARANG CITI CAPITAL LETTER ANG
118AB	᱑	WARANG CITI CAPITAL LETTER GA
118AC	᱒	WARANG CITI CAPITAL LETTER KO
118AD	᱓	WARANG CITI CAPITAL LETTER ENY
118AE	᱔	WARANG CITI CAPITAL LETTER YUJ
118AF	᱕	WARANG CITI CAPITAL LETTER UC
118B0	᱖	WARANG CITI CAPITAL LETTER ENN
118B1	᱗	WARANG CITI CAPITAL LETTER ODD
118B2	᱘	WARANG CITI CAPITAL LETTER TTE
118B3	᱙	WARANG CITI CAPITAL LETTER NUNG
118B4	ᱚ	WARANG CITI CAPITAL LETTER DA
118B5	ᱛ	WARANG CITI CAPITAL LETTER AT
118B6	ᱜ	WARANG CITI CAPITAL LETTER AM
118B7	ᱝ	WARANG CITI CAPITAL LETTER BU
118B8	ᱞ	WARANG CITI CAPITAL LETTER PU
118B9	ᱟ	WARANG CITI CAPITAL LETTER HIYO
118BA	ᱠ	WARANG CITI CAPITAL LETTER HOLO
118BB	ᱡ	WARANG CITI CAPITAL LETTER HORR
118BC	ᱢ	WARANG CITI CAPITAL LETTER HAR
118BD	ᱣ	WARANG CITI CAPITAL LETTER SSUU
118BE	ᱤ	WARANG CITI CAPITAL LETTER SII
118BF	ᱥ	WARANG CITI CAPITAL LETTER VIYO

**Lowercase vowels**

118C0	᱆	WARANG CITI SMALL LETTER NGAA
118C1	᱇	WARANG CITI SMALL LETTER A
118C2	᱈	WARANG CITI SMALL LETTER WI
118C3	᱉	WARANG CITI SMALL LETTER YU
118C4	᱊	WARANG CITI SMALL LETTER YA
118C5	᱋	WARANG CITI SMALL LETTER YO
118C6	᱌	WARANG CITI SMALL LETTER II
118C7	ᱍ	WARANG CITI SMALL LETTER UU
118C8	ᱎ	WARANG CITI SMALL LETTER E
118C9	ᱏ	WARANG CITI SMALL LETTER O

**Lowercase consonants**

118CA	᱐	WARANG CITI SMALL LETTER ANG
118CB	᱑	WARANG CITI SMALL LETTER GA
118CC	᱒	WARANG CITI SMALL LETTER KO
118CD	᱓	WARANG CITI SMALL LETTER ENY
118CE	᱔	WARANG CITI SMALL LETTER YUJ
118CF	᱕	WARANG CITI SMALL LETTER UC
118D0	᱖	WARANG CITI SMALL LETTER ENN
118D1	᱗	WARANG CITI SMALL LETTER ODD
118D2	᱘	WARANG CITI SMALL LETTER TTE
118D3	᱙	WARANG CITI SMALL LETTER NUNG
118D4	ᱚ	WARANG CITI SMALL LETTER DA
118D5	ᱛ	WARANG CITI SMALL LETTER AT
118D6	ᱜ	WARANG CITI SMALL LETTER AM
118D7	ᱝ	WARANG CITI SMALL LETTER BU
118D8	ᱞ	WARANG CITI SMALL LETTER PU
118D9	ᱟ	WARANG CITI SMALL LETTER HIYO
118DA	ᱠ	WARANG CITI SMALL LETTER HOLO
118DB	ᱡ	WARANG CITI SMALL LETTER HORR

118DC	ᱢ	WARANG CITI SMALL LETTER HAR
118DD	ᱣ	WARANG CITI SMALL LETTER SSUU
118DE	ᱤ	WARANG CITI SMALL LETTER SII
118DF	ᱥ	WARANG CITI SMALL LETTER VIYO

**Digits**

118E0	ᱦ	WARANG CITI DIGIT ZERO
118E1	ᱧ	WARANG CITI DIGIT ONE
118E2	ᱨ	WARANG CITI DIGIT TWO
118E3	ᱩ	WARANG CITI DIGIT THREE
118E4	ᱪ	WARANG CITI DIGIT FOUR
118E5	ᱫ	WARANG CITI DIGIT FIVE
118E6	ᱬ	WARANG CITI DIGIT SIX
118E7	ᱭ	WARANG CITI DIGIT SEVEN
118E8	ᱮ	WARANG CITI DIGIT EIGHT
118E9	ᱯ	WARANG CITI DIGIT NINE

**Numbers**

118EA	ᱰ	WARANG CITI NUMBER TEN
118EB	ᱱ	WARANG CITI NUMBER TWENTY
118EC	ᱲ	WARANG CITI NUMBER THIRTY
118ED	ᱳ	WARANG CITI NUMBER FORTY
118EE	ᱴ	WARANG CITI NUMBER FIFTY
118EF	ᱵ	WARANG CITI NUMBER SIXTY
118F0	ᱶ	WARANG CITI NUMBER SEVENTY
118F1	ᱷ	WARANG CITI NUMBER EIGHTY
118F2	ᱸ	WARANG CITI NUMBER NINETY

**Signs**

118FF	ᱹ	WARANG CITI OM
-------	---	----------------



## Figures.

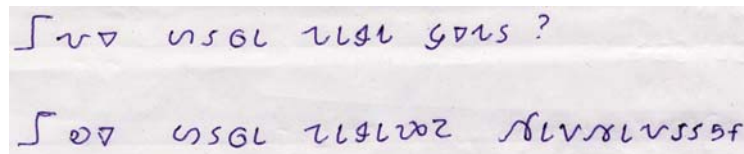



Figure 1. Sample of Warang Citi handwriting.

ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ Hbar(a)ṇ citi „Alphabet“									
Originalschr.		Transliteration		Bezeichnung der Grapheme		Phonement- sprechung			
Groß- Buchstabe	Klein- Buchstabe	Devanāgarī	Latein- Schrift	Devanāgarī	Phonetisch				
1. ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ „(einfache) Vokale“									
—	—	—	(a), (e), (o)	—	(„inhärierend“)	a ,  e ,  o			
V	v	.	M	m	ᱵᱚᱠ	ᱵᱚᱠ  m ,  ṇ			
ᱵ	ᱚ	अ, आ	A	a	ᱵᱚᱠ	ᱵᱚᱠ  a			
f	f	इ	I	i	विःइ	wiṛi, wiṛi  i			
L	l	उ	U	u	युःउ	yuṛu  u			
2. ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ sell ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ „(gemischte) Vokale“									
ᱵ	ᱚ	ए	E	e	ᱵᱚᱠ	ᱵᱚᱠ  e			
Z	z	औ	O	o	औःअ	ᱵᱚᱠ  o			
3. ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ dobri bor(o)ṇ „Ligaturen“									
Y	y	:	ṛ	ṛ	यः	'yaṛ  ṛ			
H	h	य	Y	y	यो	'yoṛ  y ,  e ,  i			
E	e	इ, ई, ऐ	I	i	ई	ṛi:  i ,  ai ,  ii			
h	h	उ, ऊ, औ	ū	ū	ऊ	ṛu:  u ,  au ,  uu			
4. ᱵᱚᱠ ᱵᱤᱨ ᱵᱚᱠ ᱵᱚᱠ ᱵᱚᱠ beyn bor(o)ṇ „Konsonanten“									
ᱵ	ᱚ	उ.	N	ṇ	ओइ	ᱵᱚᱠ  ṇ			
ᱵ	ᱚ	ग	G	g	गः	gaṛ  g			
ᱵ	ᱚ	क	K	k	कः	kaṛ  k			
ᱵ	ᱚ	ज	N	ṇ	विज	ᱵᱚᱠ  ṇ			

Figure 2a. Description of the alphabet from Pinnow 1972.



Originalschr. Groß- Klein- Buchstabe		Transliteration Devanāgarī Latein- Schrift		Bezeichnung der Grapheme Devanāgarī Phonetisch		Phonement- sprechung
E	६	ज	J j	बिज्	yü'j	j/
५	५	च	C c	बिच्	ʔüc	c/
२	३	ण	N n	त्रेण	ʔeɽ̃	n/
८	८	ड	D d	ओड्	ʔo'd	d/
L	८	ट	T t	टेःए	ʔeɽ̃	t/
१	२	न	N n	नुङ्	nuŋ	n/
७	३	द	D d	दःऽ	daʔ	d/
O	६	त	T t	औत	ʔat	t/
४	५	म	M m	ऽह्म	ʔam	m/
०	०	ब	B b	बू	bu	b/
U	३	प	P p	पुःऽ	puʔ	p/
७	५	ह	H h	ह्यी	hiyo	h/, Vokaldehnung
७	७	ल	L l	ह्ली	holo	l/
५	५	ड़	R r	हड़	hor	r/
T	४	र	R r	हर्	har	r/
४	४	ष,श	S s	षू	śu:	ś/,  s/
८	८	स	S s	सौ	sī:	s/
७०५०	हब	HB hb	—	—	—	b/,  w/

5. Zeichen für die heilige Silbe  OM

6. Zahlzeichen

१	२	३	४	५	६	७	८	९
१	२	३	४	५	६	७	८	९
१०	२०	३०	४०	५०	६०	७०	८०	९०
aber ११	१२	२३	usw.	५७	५८	५९	६०	६१

Figure 2b. Description of the alphabet from Pinnow 1972.

# 12017F



8...8...8 6f3 7203!  
 7f803 11103 9253! 0f803 9253 9253!!  
 0f803 00383 7250!!! 0f803 1126f3 7203!!!!  
 0000 0v 013 1 0 1  
 1 1 0 0 1v 0v.

120 1100 1100—1110 1110 111 110  
 12 1102 1120 123 11 0211  
 0000 0000 0110 1100 110 1210  
 0110 1102 000 1100 1111.  
 72123 1111 1111 1111 0111 11 1'  
 111211 01121 0111 0211 1111  
 00, 0111 0111 0111 0111 0111  
 111111, 1111111 111 1111 12 1 00111.  
 1111 1111111 1111 1111 1111  
 1 00111 111111 1' 11111 111111 1111

# 12017 1211

Seite 1 aus *Shar Hora* „Der Weg zum Himmel“ (Transliteration 6.1.4)

Figure 3. Sample text from Pinnow 1972.

ṁ

ṁ	ṁ	हो	हं ṁṁ
ṁ	ṁ	हो	ल ṁṁ
ṁ	ṁ	हड़	ह ṁṁ
ṁ	ṁ	हर	र ṁṁ
ṁ	ṁ	ष	ष ṁṁṁṁṁ ṁṁ
ṁ	ṁ	सी	स ṁṁṁ ṁṁ

ṁṁ ṁṁṁ  
अप पाट

ṁṁṁṁṁ ṁṁṁ  
तृतीय पाठः

ṁṁṁṁṁ ṁṁṁṁṁ  
ककर ककहारा

V— ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ,  
अं, गं, कं, जं, चं, णं, डं, टं,  
ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ,  
नं, दं, तं, मं, वं, पं, हं, लं, डं,  
ṁṁ, ṁṁ, ṁṁ.  
रं, षं, सं।

ṁ— ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ,  
गा, का, जा, चा, णा, डा, टा, ना,  
ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ,  
दा, ता, मा, वा, पा, हा, ला, डा, रा,  
ṁṁ, ṁṁ,  
षा, सा।

ṁ— ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ, ṁṁ,  
गि, कि, जि, चि, णि, डि, टि, नि, दि, ति,

Figure 4a. Sample text with syllabary from Pinnow 1972.

4

8

५f, ०f, १f, ७f, ७f, ५f, ४f, ४f, ५f.  
मि, वि, पि, हि, लि, डि, रि, षि, सि।

L— १६, १६, ६६, ६६, १६, १६, १६, १६, १६,  
गु, कु, गु, जु, चु, गु, डु, टु, नु, दु,  
६६, १६, ०६, १६, ७६, ७६, ५६, ४६, ४६,  
तु, मु, वु, पु, हु, लु, डु, रु, षु, सु।

V— १७, १७, ६७, ६७, १७, १७, १७, १७, १७,  
गो, के, जे, जे, चे, णे, डे, टे, ने, दे,  
६७, १७, ०७, १७, ७७, ७७, ५७, ४७, ४७,  
ते, मे, वे, पे, हे, ले, डे, रे, षे, से।

Z— १८, १८, ६८, ६८, १८, १८, १८, १८, १८,  
गो, को, जो, जो, चो, णो, डो, टो, नो, दो,  
६८, १८, ०८, १८, ७८, ७८, ५८, ४८, ४८,  
तो, मो, वो, पो, हो, लो, डो, रो, षो, सो।

Y— १५, १५, ६५, ६५, १५, १५, १५, १५, १५,  
गः, कः, जः, जः, चः, णः, डः, टः, नः,  
१५, ६५, १५, ०५, १५, ७५, ७५, ५५, ४५,  
दः, तः, मः, वः, पः, हः, लः, डः, रः,  
४५, ५५.  
षः, सः।

H— ११, ११, ६१, ६१, ११, ११, ११, ११, ११,  
गै, कै, जै, जै, चै, णै, डै, टै, नै,  
११, ६१, ११, ०१, ११, ७१, ७१, ५१, ४१,  
दै, तै, मै, वै, पै, है, लै, डै, रै,  
४१, ५१.  
षै, सै।

Seite 4 aus Ho H(a)y(a)m P(a)ham Puti

Figure 4b. Sample text with syllabary from Pinnow 1972.







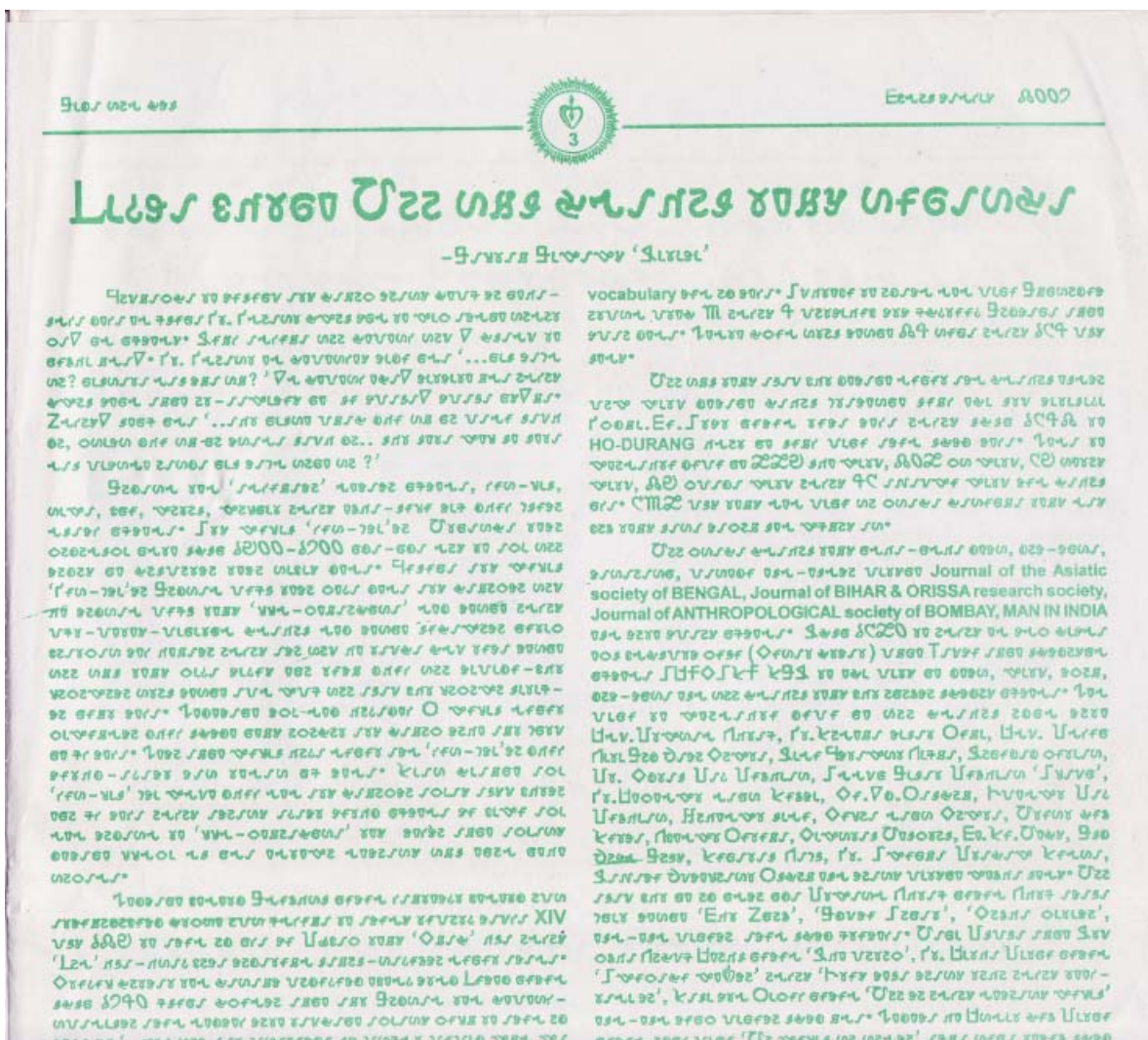


Figure 7. Part of a newsletter in Warang Citi.

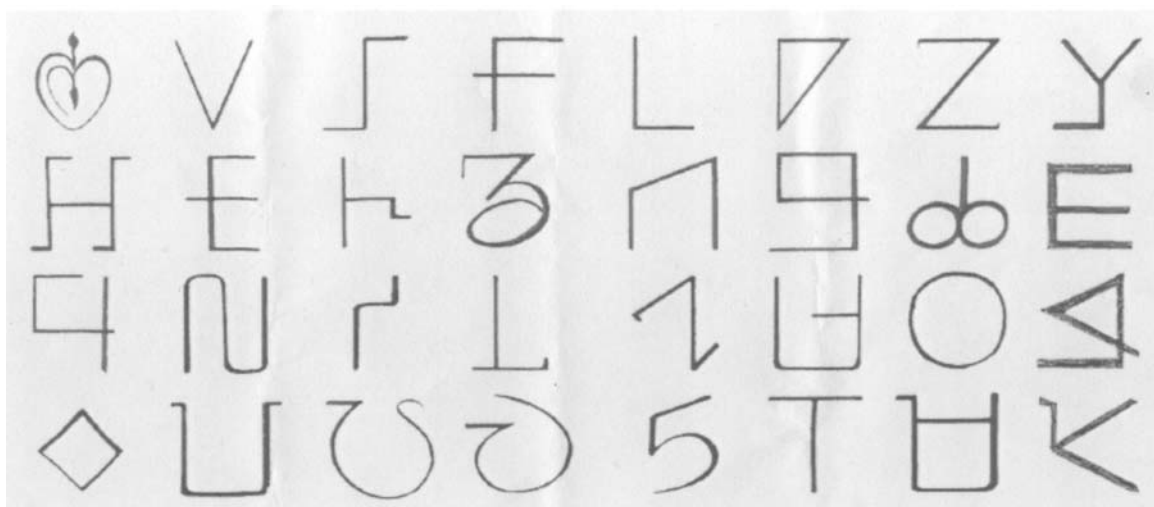
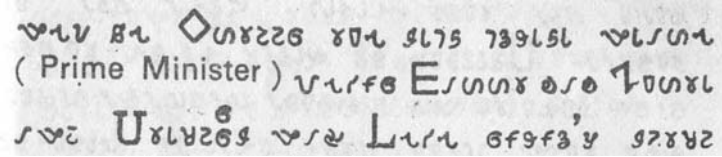


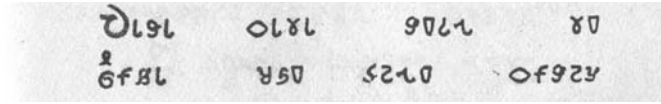
Figure 8. Warang Citi alphabet as written out by K. C. Naik.



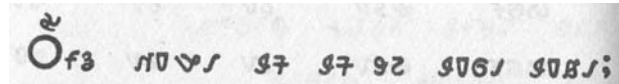




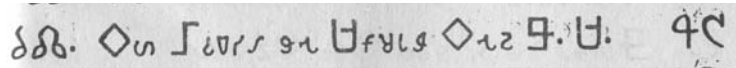
**Figure 16.** Text showing Uṣṣṭama /puru sutam/, with a ṣṭ conjunct.



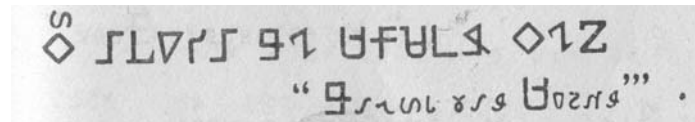
**Figure 17.** Text showing triyu /tiryu/, with a ṭri conjunct.



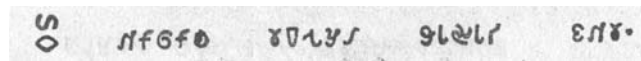
**Figure 18.** Text showing Tsina /tisij/, with a ṭsi conjunct.



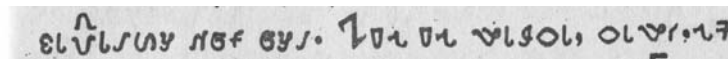
**Figure 19.** Text showing Bha~Baha /ba:/, without a ṭ conjunct.



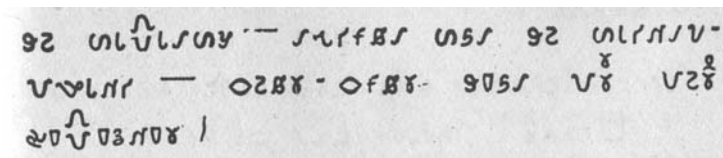
**Figure 20.** Text showing Bha /ba:/, with a ṭba conjunct.



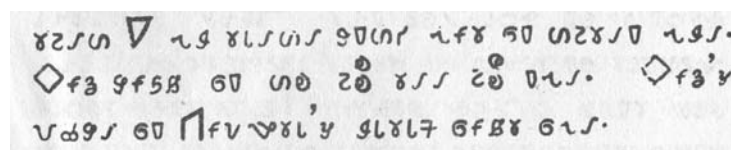
**Figure 21.** Text showing bha /ba:/, with a ṭba conjunct.



**Figure 22.** Text showing juppuāhṭ, (juppuāḥṭ) /jupuwaḥ/, with a ṭpu conjunct.



**Figure 23.** Text showing ṭpa parra /par/ and ṭpo porra /por/, each with a ṭ conjunct. It appears that the right-side-up conjunct form on ṭpa parra is an error and should be ṭpa parra.



**Figure 24.** Text showing ṭla olla /ol/ and ṭlo olla, each with a ṭ conjunct. It appears that the right-side-up conjunct form on ṭla olla is an error and should be ṭla olla.

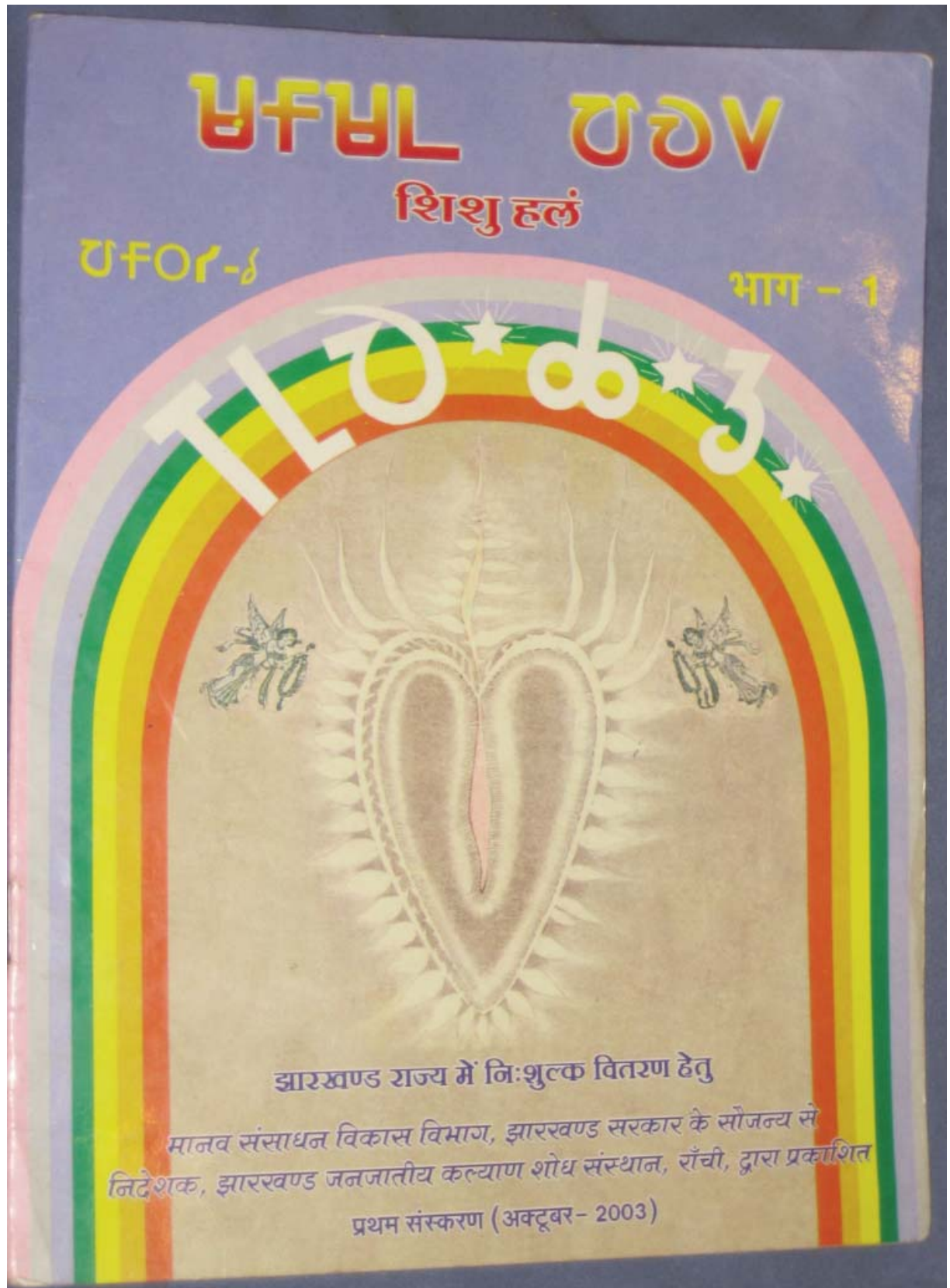


Figure 25. The cover of UFUL UDU (Shisu halam).

## A. Administrative

1. Title

**Revised proposal for encoding the Warang Citi script in the SMP of the UCS**

2. Requester's name

**UC Berkeley Script Encoding Initiative (Universal Scripts Project)**

3. Requester type (Member body/Liaison/Individual contribution)

**Liaison contribution.**

4. Submission date

**2012-04-19**

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

**Yes.**

6b. More information will be provided later

**No.**

## B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

**Yes.**

1b. Proposed name of script

**Warang Citi.**

1c. The proposal is for addition of character(s) to an existing block

**No.**

1d. Name of the existing block

2. Number of characters in proposal

**84.**

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

**Category A.**

4a. Is a repertoire including character names provided?

**Yes.**

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

**Yes.**

4c. Are the character shapes attached in a legible form suitable for review?

**Yes.**

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

**Michael Everson.**

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

**Michael Everson, Fontographer.**

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

**Yes.**

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

**Yes.**

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

**Yes.**

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

**See above.**

## C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

**Yes, in N1958, N3411, N3668.**

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

**Yes.**

2b. If YES, with whom?

**Dobro Buriuli (teacher of Warang Citi), Chandra Mohan Haibru, K.C. Naik Biruli, Anna Pucilowski, K. David Harrison, Gregory Anderson.**

2c. If YES, available relevant documents



3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

**Speakers of the Ho language, whose population is 1,026,000, with 25%-50% literacy, according to the SIL *Ethnologue*. The *Ethnologue* mentions only the Oriya and Devanagari script areas.**

4a. The context of use for the proposed characters (type of use; common or rare)

**To write the Ho language.**

4b. Reference

5a. Are the proposed characters in current use by the user community?

**Yes.**

5b. If YES, where?

**In India.**

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

**Yes.**

6b. If YES, is a rationale provided?

**Yes.**

6c. If YES, reference

**Contemporary use and accordance with the Roadmap.**

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

**Yes.**

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

**No.**

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

**No.**

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

**No.**

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences?

**No.**

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

**No.**

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

**No.**

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

**No.**

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?